

**Amendments to the Specification:**

Please replace paragraph page 9 line 21 to page 10 line 1 with the following amended paragraph:

[Non-patent document 1]

WTP (WebSphere Transcoding Publisher,  
<http://www-6.ibm.com/jp/software/network/transcoding/>)

Please replace paragraph page 10 lines 2-7 with the following amended paragraph:

[Non-patent document 2]

CHIP[I] Ito "Construction method of distributed applications by integration of GUI parts and WEB services," Japan Society for Software Science and Technology WISS 2001 Proceedings  
(<http://ea.memo.hokudai.ac.jp/people/ito/CHIP/index.J.html>)

Please replace paragraph page 10 lines 8-10 with the following amended paragraph:

[Non-patent document 3]

IBM mySite Outliner  
(<http://www-6.ibm.com/jp/pc/clubibm/msol/index.shtml>)

Please replace paragraph page 10 lines 11,12 with the following amended paragraph:

[Non-patent document 4]

DiffWeb (~~<http://www.diffweb.com/>~~)

Please replace paragraph page 10 lines 13 and 14 with the following amended paragraph:

[Non-patent document 5]

HTML Diff (~~<http://www-db.stanford.edu/e3/e3.html>~~)

Please replace paragraph page 10 lines 15 and 16 with the following amended paragraph:

[Non-patent document 6]

MindIt (~~<http://mindit.net/mind.com/mindit.shtml>~~)

Please replace paragraph page 25, lines 14-20 with the following amended paragraph:

For example, the next URLs are listed as adjacent structured/hierarchical contents to a Web content as the target content of which URL is

~~<http://www.asahi.com/0606/news/national06015.html>~~

~~<http://www.asahi.com/0606/news/national06012.html>~~

~~<http://www.asahi.com/0606/news/national06013.html>~~

~~<http://www.asahi.com/0606/news/national06014.html>~~

Please replace paragraph page 26 lines 5-18 with the following amended paragraph:

For example, the next URLs are listed as adjacent structured/hierarchical contents to a Web content as the target content of which URL is

[http://www.cnn.com/2000/US/06/05/  
sea.based.defense/index.html](http://www.cnn.com/2000/US/06/05/seabased.defense/index.html):  
<http://www.cnn.com/2000/US/06/05/dday.remembrance/index.htm>1  
[http://www.cnn.com/2000/US/06/05/helicopter.escape.03/index.  
html](http://www.cnn.com/2000/US/06/05/helicopter.escape.03/index.html)  
[http://www.cnn.com/2000/US/06/05/curbing.terrorism.02/index.  
html](http://www.cnn.com/2000/US/06/05/curbing.terrorism.02/index.html)

Please replace paragraph page 38 line 9 to page 40 line 4 with the following amended paragraph:

<sup>66</sup><http://www.asahi.com/business/update/1019/002.html>

In order to generate an appropriate matching pattern even in the case as described above, the present invention introduces a concept titled "adjacent structured/hierarchical contents to a target content." The adjacent structured/hierarchical contents are structured/hierarchical contents, which have URLs analogous to that of the target content and are made to belong to the same group as that of the target content in the case of a matching determination by means of the matching pattern. The analogous range of the URLs is varied depending on the extent to which the author determines that differences of the structured/hierarchical contents are allowable and the different contents belong to the same group. The URLs include directories (portions partitioned by forward slashes in the example of the business article in Asahi Shimbun) in the respective hierarchies. When the URLs of the contents to be determined whether or not they are the adjacent structured/hierarchical contents are collated with the URL of the target content, if directories up to a

predetermined number (one or more) of hierarchies from the uppermost hierarchy are identical and only directories in lower hierarchies from the hierarchies where the directories are identical, the content portions subjected to the determination may be determined as adjacent content portions. Specific examples of the adjacent content portions are listed as follows. In the next cases, the structured/hierarchical contents subjected to the determination are determined as the adjacent structured/hierarchical contents. (a) Only a portion recognized as a date in the URL differs from that of the target content. In the above-described example of the business article of Asahi Shimbun, the relevant portion is "1019." (b) Only a portion used as numbering in the URL differs from that of the target content. In the above-described example of the business article of Asahi Shimbun, the relevant portion is "002.html." (c) Only the above-described (a) and (b) differ from those of the target content.

Please replace paragraph page 77 lines 14 to 20 with the following amended paragraph:

The case where the past page is not present occurs not only when the caching of the past pages is not performed but also when the URLs are generated every day. For example, in the case where a date is utilized as a part of a URL as in a URL of a newspaper article, it is obvious that no past pages can be present  
(<http://www.asahi.com/international/update/1005/010.html>).

Please delete paragraph page 78 lines 5-7 beginning with "Example:".

Please delete paragraph page 78 lines 8-9 beginning with "Adjacent URL:".

Please replace the paragraph page 96 line 11 to page 99 line 2 with the following paragraph:

The following sample is premised on the implementation as described above. Note that the following is a description example according to the relaxNG format.

```
<?xml version="1.0" ?>
<grammar xmlns="http://relaxng.org/ns/structure/0.9">
  <start>
    <element name="table">
      <attribute name="width">
        <value>168</value>
      </attribute>
      <zeroOrMore>
        <choice>
          <ref name=lfreeAttributesTABLE"/>
        </choice>
      </zeroOrMore>
      <element name="tbody">
        <zeroOrMore>
          <choice>
            <ref name=lfreeAttributesTBODY"/>
          </choice>
        </zeroOrMore>
        <oneOrMore>
          <element name="tr">
            <attribute name="bgcolor">
              <value>ffffff</value>
            </attribute>
```

```
<zeroOrMore>
<choice>
<ref name="freeAttributeTR"/>
</choice>
</zeroOrMore>
<element name="td">
<zeroOrMore>
<choice>
<ref name=lfreeAttributesTD"/>
</choice>
</zeroOrMore>
<element name="small">
<zeroOrMore>
<choice>
<ref name=lfreeAttributesSMALL"/>
</choice>
</zeroOrMore>
<element name="a">
<zeroOrMore>
<choice>
<ref name=lfreeAttributesA"/>
</choice>
</zeroOrMore>
<text/>
</element>
</element>
</element>
</element>
</oneOrMore>
```

</element>

</element>

</start>

Please replace the paragraph at page 102 line 8 to page 103 line 20 with the following amended paragraph:

```
<?xml version="1.0" encoding="utf-8" ?>
<rdf :RDF
xmlns="http://purl.org/rss/1.0/"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xm: lang="ja">
<channel rdf:about="http://news.lycos.co.jp/topics/rss.rdf">
<title>News LYCOS Saishin Topics</title>
<link>http://news.lycos.co.jp/topics</link>
<items>
<rdf:Seq>
<rdf:li
rdf:resource="http://news.lycos.co.jp/topics"/>
</rdf:Seq>
</items>
</channel>
<item
rdf:about="http://news.lycos.co.jp/topics/society/maff.
html">
<title>isahayawan kinpaku-no-naka koji saikai</title>
<link>http://news.lycos.co.jp/topics/society/maff.html
</link>
</item>
```

```
<item rdf:about="http://news.lycos.co.jp/topics/world/operation.html">
<title>arukaide sento-in amerika-ni-tokoh</title>
<link>http://news.lycos.co.jp/topics/world/operation.html
</link>
</item>

<item rdf:about="http://news.lycos.co.jp/topics/computer/ms.html">
<title>maikurosofuto kadenbunya shinshutsu</title>
<link>http://news.lycos.co.jp/topics/computer/ms.html
</link>
</item>
```

Please replace the paragraph page 110 line 6 to page 111 line 10 with the following paragraph:

[Other Example 4: Application to Information Aggregator]  
Partial cut out of Web pages and integration of information are broadly performed in a portal construction system such as the IBM Portal Server and an information extraction/submission system such as the IBM mySiteOutliner. The present invention is applicable to these systems. For example, in the IBM mySiteOutliner, XPath as below is held in a definition file in order to extract a headline link list from the Web page.

```
<ClippingDefinition>
<id>2</id>
<links>
<link title="Club IBM Top
Page> http://www.ibm.com/jp/pc/clubibm/index.html</link>
</links>
```



```
<urldata>
<url source="Club IBM"> http://www.ibm.com/jp/pc/
    clubibm/index.html</url>
<xpathlists>
xpath name=l/body textl/>
/html[1]/body[1]/table[2]/tbody[1]/tr[1]/td[2]/table[2]/tbody[1]/tr[5]/td[2]/table[1]/tbody[1]/tr[1]/td[1]/table[2]/tbody[1]/tr[2]/td[1]
</xpath>
</xpathlists>
</urldata>
</ClippingDefinition>
```